

Contrast in Terahertz Images of Archival Documents—Part I: Influence of the Optical Parameters from the Ink and Support

Tiphaine Bardon¹ • Robert K. May² •
J. Bianca Jackson³ • Gabriëlle Beentjes⁴ •
Gerrit de Bruin⁴ • Philip F. Taday² • Matija Strlič¹

Received: 11 September 2016 / Accepted: 19 December 2016 /

Published online: 17 January 2017

© The Author(s) 2017. This article is published with open access at Springerlink.com

Abstract This study aims to objectively inform curators when terahertz time-domain (TD) imaging set in reflection mode is likely to give well-contrasted images of inscriptions in a complex archival document and is a useful non-invasive alternative to current digitisation processes. To this end, the dispersive refractive indices and absorption coefficients from various archival materials are assessed and their influence on contrast in terahertz images from historical documents is explored. Sepia ink and inks produced with bistre or verdigris mixed with a solution of Arabic gum or rabbit skin glue are unlikely to lead to well-contrasted images. However, dispersions of bone black, ivory black, iron gall ink, malachite, lapis lazuli, minium and vermilion are likely to lead to well-contrasted images. Inscriptions written with lamp black, carbon black and graphite give the best imaging results. The characteristic spectral signatures from iron gall ink, minium and vermilion pellets between 5 and 100 cm^{-1} relate to a ringing effect at late collection times in TD waveforms transmitted through these pellets. The same ringing effect can be probed in waveforms reflected from iron gall, minium and vermilion ink deposits at the surface of a document. Since TD waveforms collected for each scanning pixel can be Fourier-

Electronic supplementary material The online version of this article (doi:10.1007/s10762-016-0351-0) contains supplementary material, which is available to authorized users.

✉ Matija Strlič
m.strlic@ucl.ac.uk

¹ UCL Institute for Sustainable Heritage, The Bartlett, UCL Faculty of the Built Environment, London WC1H 0NN, UK

² TeraView Ltd, Platinum Building, St John's Innovation Park, Cambridge CB4 0DS, UK

³ Department of Physics, Capital Normal University, Beijing 100010, China

⁴ National Archives of The Netherlands, P.O. Box 90520, 2509 LM The Hague, The Netherlands

transformed into spectral information, terahertz TD imaging in reflection mode can serve as a hyperspectral imaging tool. However, chemical recognition and mapping of the ink is currently limited by the fact that the morphology of the document influences more the terahertz spectral response of the document than the resonant behaviour of the ink.

Keywords Terahertz · Time-domain · Imaging · Documents · Inks · Spectroscopy

1 Introduction

Digitisation of records ensures and improves accessibility to archival records, helping archival institutions in their legal obligation [1–4]. Current digitisation of textual archives is mostly performed by overhead cameras, scanners with a flat scanning bed or planetary book scanners [5, 6]. While these technologies are adequate for scanning flat loose sheets in a non-invasive way, their use to digitise bound documents may cause mechanical stress to the document and increases risks of damage. In some cases, the decision to unbind or cut the bound document to flatten it and facilitate digitisation might be taken [7]. The most optimal digitisation option adapted to very fragile bound or closed documents (i.e. books or letters) would be an in-depth scanning tool which could produce an image of each page of the entire document in one single non-invasive scan while the document remains closed. Terahertz time-domain (TD) technology is a non-invasive depth-imaging technology which first emerged in the mid-1990s [8], but its use is now increasing through its commercialisation. Studies using models have shown that this technology can uncover textual information not readily accessible to the eye [9–15]. Yet, many of these models were made with graphite (lead) pencil or “carbon black” ink to write the inscriptions, materials which are known to be highly reflective and absorbent in the terahertz region. As for the other models, no information is provided on the nature of the ink in the pen used for the inscriptions or of the paint used to cover the inscriptions. Wallace et al. [10] attribute the legibility of ink inscriptions within a folded letter in an envelope to an increase in absorption and/or refractive index caused by the presence of the ink while admitting that the refractive index of the specific ink they used is unknown to them. Mittleman et al. [13], however, stipulate that inscriptions written with “conventional ink” cannot be detected with terahertz imaging as dye molecules have low absorption in the terahertz range. Besides, inscriptions in some historical documents investigated by the authors could not be revealed with terahertz TD technology. There therefore seems to be a need for a better understanding of what makes inscriptions legible in terahertz images. Specifically, since terahertz imaging of a valuable and fragile historical document would either imply risking the transportation of the document to one of the few research facilities equipped with a terahertz imaging set-up, or purchasing a terahertz set-up for in situ analysis and recruiting staff with expertise in terahertz imaging, it is crucial to determine beforehand the likelihood of success of the scan based on the optical properties of the ink material in the terahertz region. If this likelihood is assessed to be low, another imaging technology might be considered (e.g. X-ray attenuation imaging or edge-illumination X-ray phase contrast imaging, combined with computed tomography).

In addition, most of the studies mentioned above used a terahertz TD imaging instrument set up in transmission mode [9, 12–14]. Such configuration can be valuable to image single-sheet documents with a censored text covered with paint. However, for stratified documents with inscriptions present on several layers, such as letters in envelopes, leaflets or books, the signal transmitted through the document would undergo absorption from all inscriptions across

its path. As a result, the textual information from a specific layer would be difficult to recover on the image obtained from the amplitude of the transmitted pulse, as shadows of inscriptions from other layers within the document would be superimposed to the information of interest [9]. In reflection mode, however, such shadows or occlusions on the image obtained from the amplitude of the pulse reflected on the n^{th} layer can be removed by using the image obtained from the amplitude of the pulse reflected on the $(n-1)^{\text{th}}$ layer. The reflection configuration therefore seems more appropriate to image archival documents.

Finally, some pigments have been shown to possess a dispersive behaviour: their terahertz spectral signatures can be recovered from the Fourier-transform of TD waveforms [16, 17]. Terahertz TD imaging can therefore also serve as a hyperspectral imaging tool. Combined with a highly sensitive detection system, it has been used in transmission mode as a chemical imaging tool to discriminate ink inscriptions from stains on the document [16]. Whether the reflection mode can be used for this purpose is yet to be determined. The use of terahertz TD imaging for chemical imaging is relevant to paper and book conservators as it could help distinguishing between inert and corrosive inks, such as carbon-based and iron gall inks [18], in a non-invasive way, and treat or store the document accordingly.

The aims of this study are to assess the optical properties in the terahertz region 0.15–3 THz from materials commonly found in archives and to use this information to better understand and interpret contrast in terahertz images from ink inscriptions scanned in reflection mode. This study further exploits these terahertz optical properties to refine the use of terahertz TD technology for chemical mapping.

2 Materials and Methods

2.1 Pigments and Binders

The selection of pigments and inks used in this study was based on the frequency of their use in archival documents [19–21], their predisposition to degrade the support, the availability of their components and the feasibility of their production in a laboratory. The nature of pigments acquired from art suppliers was confirmed with Fourier-transform infrared (FTIR) spectroscopy, acquired either in attenuated total reflectance (ATR) or transmission mode. Measurements were performed using the Alpha-P Bruker FTIR spectrometer with OPUS software, version 6.5. Spectroscopic data was acquired within the wavenumber range 375–4000 cm^{-1} , where characteristic absorption peaks from both organic and inorganic compounds can be found. Sixty-four scans were performed at 4 cm^{-1} spectral resolution. These spectra are supplied in Online Resource 1.

Lamp black (Product ID: LC17033F), carbon black (LC17073F), graphite (LC17004H), bone black (LC17023F), ivory black (LC20011C), malachite (LC20132E), lapis lazuli (LC20002E), vermilion (LC12181D), minium (LC12053F), verdigris (LC20091C) and rabbit skin glue (LC27484H) powders were acquired from Cornelissen and Son, London, while bistre (12100) and Arabic gum (63300) were acquired from Kremer Pigmente, Munich. Dry binders such as rabbit skin glue and Arabic gum were ground into powder in a coffee grinder. Sepia ink was made by diluting 4 g of cuttlefish ink (acquired from a local fishmonger) in 10 mL of deionised water. Iron gall ink was prepared following the composition of “IG1” described elsewhere [18].

It should be noted that ivory black was acquired as “genuine ivory black”, while bone black was acquired from the same supplier as “ivory black”. A discussion with the supplier revealed that bone black was more likely to have been produced from calcined bones than ivory. This was confirmed by the presence of vibrations attributed to cyano stretches in cyanamide, characteristic of burnt bone, at 697 and 2012 cm^{-1} [22, 23] in the ATR-FTIR spectrum of the “ivory black” supply, while these bands were not present in the spectrum of the “genuine ivory black” supply. Furthermore, neither the ATR nor transmission FTIR spectra from vermilion showed spectral feature, which is consistent with the literature [23, 24]. However, Derrick et al. remind us that the red pigments cadmium red and realgar also show no spectral band in the mid-infrared region [23]. Still, the terahertz spectral signatures of these three red pigments are different [17]: vermilion shows sharp spectral bands around 1.1, 1.3 and 2.6 THz, which is in agreement with the terahertz spectrum of the vermilion supply presented in Fig. 7. It can therefore be concluded that the “vermilion” supply is constituted of pure vermilion. The “minium” supply showed intense spectral features at 387, 453 and 532 cm^{-1} , as previously reported for minium and lead(II, IV) oxide (Pb_3O_4) [25, 26]. It also contains minute amounts of lead carbonates, as attested by the weak bands at 685 and 1410 cm^{-1} [27, 28]. The broad and weak absorption bands at 489 and 511 cm^{-1} may attest the presence of litharge and massicot [29], respectively, two polymorphs of lead(II) oxide (PbO) [20]. The supplier of lapis lazuli was not able to attest the provenance of this pigment (Chilean or Afghan). Yet, studies report that Afghan lapis lazuli can be differentiated from Chilean lapis lazuli and synthetic ultramarine thanks to the presence of an absorption band at approximately 2340 cm^{-1} [23, 30, 31]. This vibrational band was observed in none of the three ATR-FTIR spectra from the lapis lazuli powder investigated here: it can therefore be deduced that our lapis lazuli supply is not of Afghan origin. Its ATR-FTIR spectrum showed intense overlapping bands at 960–1000 cm^{-1} also found in the lazurite sample from the Forbes Pigment Database [32], but shows no band associated to calcium carbonate (RRUFF ID: R040070) or pyrite (RRUFF ID: R050190) from the RRUFF online database [33]. It can therefore be assumed that the supply is largely composed of pure lazurite. Finally, FTIR spectra from the supplies of bistre, malachite, verdigris, rabbit skin glue and Arabic gum confirmed that they are respectively constituted of chars from burnt lignin (a polymeric aromatic organic substance in wood) [34], copper carbonate ($\text{Cu}_2\text{CO}_3(\text{OH})_2$) (RRUFF ID: R050508) [33, 35], copper(II) acetate monohydrate ($\text{Cu}(\text{CH}_3\text{COO})_2 \cdot \text{H}_2\text{O}$) [27, 36], proteins [23] and polysaccharides also found in Arabic gum of industrial grade [23, 37].

2.2 Pellet Preparation

For terahertz spectroscopic analysis, dry powders were analysed, and the solutions of sepia ink and iron gall inks were therefore left to dry on a ceramic plate, scraped off and ground into a powder with a pestle and mortar before analysis. Pure powders of binders, inks or pigments were pressed into three 400-mg pellets with a 10% mass concentration of sample to high density polyethylene (HDPE) powder and a diameter of 13 mm, following the optimal procedure described elsewhere [38].

2.3 Supports

Historically, cotton and linen rags have been widely used as a source of cellulosic fibrous materials for paper making. However, paper sheets containing only cellulosic fibres are

water absorbent: aqueous inks would diffuse into the porous fibre mesh and the resulting ink lines would be less defined. Most paper sheets are therefore sized, i.e. impregnated with substances to prevent ink from spreading into paper fibres [39, 40]. Paper sheets can be heavily sized, lightly sized or not sized at all [40]. Parchment is prepared from wet un-haired animal skin soaked in a lime solution ($\text{Ca}(\text{OH})_2$). After the removal of hair from the limed skin, the drying stretched skin is rubbed with a pumice stone to smoothen its surface. To improve the smoothness and suitability of the parchment skin for writing purposes, its surface can further be coated with egg white, oil, glue or a solution made of soaked parchment scraps, or can be treated with hot water to form a superficial layer of size [41–43]. This process further ensured that the writing ink would not penetrate too deeply in the parchment fibre mesh. To improve the adhesion of ink on parchment and prevent it from running, the surface may be dusted with powdered chalk (CaCO_3), calcium sulphate dihydrate ($\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$) and quicklime (calcium oxide, CaO) [44]. Six rag paper sheets from the nineteenth century and one parchment sheet from 1752, from the Institute for Sustainable Heritage Historic Reference Material Collection, were used as supports. To observe the influence of paper coating on terahertz images, three paper sheets were additionally sized by immersing them in a gelatine solution for less than 10 s, then placing them between two sheets of polyester and pressing them between wooden boards. The preparation of the size followed the instructions from Kolar and Strlic [45], mixing 12,500 mg of gelatine from bovine skin (Product number: G9382, Sigma Aldrich) and 1425 mg of alum (aluminium sulphate hydrate, Product number: 368458, Sigma Aldrich) in 0.5 L of deionised water. Areas of $2 \times 2 \text{ cm}^2$ were cut from each sheet of rag paper and from three non-adjacent areas of the parchment sheet for terahertz spectroscopic analysis.

2.4 Terahertz Time-Domain Spectroscopy

Pellets and supports were analysed in transmission mode using the TPS Spectra 3000 (TeraView Ltd., Cambridge, UK). Samples were placed in a nitrogen-purged sample compartment, at the focal plane of a focused terahertz beam; 1800 scans were acquired at a scan rate of 30 Hz and co-averaged to minimise random noise. The working spectral range of the frequency-domain (FD) signals obtained through the Fourier-transform of the waveforms was $5\text{--}100 \text{ cm}^{-1}$, with a spectral resolution of 1.2 cm^{-1} . The signal transmitted through a 360-mg pellet of pure HDPE powder was used as a reference measurement for the analysis of pellets, while the signal transmitted through the empty sample compartment was used as a reference for the analysis of supports. Constant and frequency-dependent refractive indices, together with frequency-dependent absorption coefficients, were calculated as described elsewhere [38]. Additional pulses resulting from multiple internal reflections on the sheet surfaces were detected in the waveform transmitted through each sheet of support. Such phenomenon is often called a Fabry-Perot or “etalon” effect [46]. These subsequent pulses overlapped with the main pulse transmitted through the sheet and therefore changed the time delay between the main peak of the raw waveform transmitted through the sample sheet and the main peak of the reference waveform. To avoid such errors in readings of the time delay, also inducing errors in the calculation of the constant refractive index, deconvolution [11, 47] was used to separate the main transmitted pulse from the pulses originating from the etalon effect. A bespoke deconvolution algorithm with a double Gaussian filter function was implemented in Matlab [48] for this purpose.

2.5 Ink Preparation and Application

For terahertz imaging experiments, ink solutions were prepared by dispersing the pigment, with no further grinding, in a solution of Arabic gum. Although a broad range of ratios of binder over water can be found in the literature, the same proportion recommended by Kolar and Strlic [45] was consistently used (i.e. 625 mg of Arabic gum with 25 mL of deionised water heated at 70 °C and left to cool down before the addition of the pigment), as it gives suitable viscosity and increases control over the experiments. Inks were prepared with a mass ratio of pigment to gum of 1:2. Mass ratios were preferred over volume ratios, as the volume of a pigment would be difficult to measure accurately due to its inter- and intra-particle pores, and as the penetration properties of the binder into the pigment pores are not known. To demonstrate the influence of ink on image contrast in a document, any indentations left by the writing tool should be avoided, as they can influence the image contrast [49]. Application with a brush or using screen printing was also avoided as the volume of ink and pigment concentration applied could not be uniformly distributed, which would have increased the complexity of the experiment and interpretation. Application with an ink jet was not considered, due to the colloidal nature of the pigment suspensions, which would have risked clogging the ink jet heads. For these reasons, 2 mL of each fresh ink was applied on historical rag paper and parchment sheets using pipettes. The resulting ink deposits had a diameter of 2 ± 0.5 mm, depending on the affinity of the ink to the support. Although the diameter of ink inscriptions written with most contemporary pens is usually smaller than 2 mm, some inscriptions written with quills, nib pens or brushes can be wider than 2 mm: the diameter of these ink droplets is therefore not unrealistic with regards to the width of archival ink inscriptions. Additionally, the ink spot was significantly larger than the step size and associated spatial resolution of the imaging set-up (see section below), which ensured that observation of the dot in the terahertz image was possible.

Despite the care taken in controlling the amount of pigment and ink applied on the support, it should be noted that the colloidal nature of these ink suspensions often led to a non-uniform deposition of particles during evaporation of water from the ink drop, creating ring-shaped solid residues (also known as “coffee ring effect” [50]). Additionally, application of ink with a pipette sometimes introduced microscopic bubbles in the droplet, creating local depletion of particle concentration. Finally, ink may penetrate non-uniformly into fibres constituting the support. All these phenomena led to the non-uniform distribution of ink thickness and pigment concentration within some ink dots: these variations of pigment concentration and topography have been recorded by acquiring micro-photographs of the ink dots (displayed alongside terahertz images of the same ink dots in Figs. 2, 3, 4 and 5) and accounted for in the interpretation of terahertz images.

2.6 Terahertz Time-Domain Imaging

2.6.1 Set-up

Terahertz images were acquired in reflection mode by fitting a reflection imaging module in the nitrogen-purged sample compartment of TPS Spectra 3000. This module comprises of a software-controlled motorised scanning table with a set of ellipsoidal mirrors underneath, focusing the generated terahertz pulses onto the holed scanning stage at an angle of incidence of 30°. Samples were placed on the scanning table and maintained steady and flat by placing them between two ring-shaped magnets taped to the stage.

2.6.2 Image Resolution

We define spatial resolution as the minimum distance between two separate features on a captured object leading to discernible features in the image of this object. Theoretical criteria, such as the Rayleigh or the Buxton criteria, are commonly applied to estimate the spatial resolution of an optical system [51, 52]. These criteria use the shape of the 2D lateral distribution of the irradiance or amplitude of the electromagnetic field emitted by a point source, also called the point spread function (PSF). For example, the Rayleigh criterion assumes that the PSF of the irradiance is a first-order Bessel function in order to evaluate the limit of resolution $r = (1.22 * f * \lambda) / D$, radius of the first dark Airy ring of the PSF. The optics of the terahertz imaging system used in this study are such that the ratio of the focal length f and aperture diameter D of the imaging system is equal to 1. It would follow that the limit of resolution is approximately 180 μm at 2 THz. However, to confirm the distribution governing the PSF of the amplitude of the terahertz pulsed electromagnetic field and better select the relevant criterion for spatial resolution, the emission profile from the terahertz source would need to be recorded at the focal plane of the imaging system. While placing a CCD camera at the focal plane would only enable to obtain an image of the irradiance profile (and not the amplitude of the pulse at a specific time or frequency) [53], an alternative is to place a terahertz detector on a translation stage at the focal plane to record the terahertz pulse point by point along a line [54]. The subsequent Fourier-transform of each of these pulses recorded at different positions would enable to recover the discrete PSF of the terahertz electromagnetic field amplitude at specific frequencies. However, since a commercial terahertz set-up with fixed optics and antennas was used in our study, such experiment could not be performed. Another alternative consists in recording the minimum distance between two separate features on a 1951 USAF resolution test target (Product ID: R3L3S1P and R3L3S1N, Thorlabs Inc) leading to discernible features in the image of this test target. The amplitude at 2 THz of the Fourier-transformed reflected time domain waveform recorded at each location in a regularly spaced 2D grid was used to produce a 2D image of the target. At the limit of resolution defined by Rayleigh, two features on the grid are not resolved if their composite intensity distribution, sum of the PSF from the two features, exhibits a central depression corresponding to a decrease in intensity of 19% [51]. For the image of the grid at 2 THz acquired with a lateral step size of 100 μm , such central depression is observed for features spaced by 180 μm . It is interesting to note that this figure coincides with the diffraction limit calculated above using the Rayleigh criterion. The following assumptions may therefore hold true in our case: (i) the PSF from the pulsed terahertz beam follows a first-order Bessel distribution, and (ii) the resolution of the system is only limited by the diffraction phenomenon when using a small scanning step size. Using the maximum amplitude of the reflected time domain waveform to produce an image of the grid, the limit of resolution as defined by a central depression in amplitude of 19% was observed for features spaced by 280 μm for a lateral step size of 100 μm . It should finally be reminded that this limit of 19% was considered satisfactory by Lord Rayleigh because it matched what the eye could detect through telescopes or microscopes [55]. Considering the resolution and size of today's computer monitors, this limit might be obsolete and might be better defined as the spacing at which a central depression

similar to the standard deviation of the amplitudes collected from a blank area of the grid is reached. On the image of the grid produced from the maximum amplitude of the pulse, this new limit is reached for features spaced by 180–200 μm , for a lateral step size of 100 μm ,

2.6.3 Signal to Noise Ratio

The signal to noise ratio is described here as the maximum value of the dynamic range. The dynamic range was calculated using the logarithmic ratio of the Fourier-transform of the TD signal reflected from a gold-coated mirror and the frequency-domain noise floor [56, 57]. The signal to noise ratio was measured to be ≈ 54 dB at 0.6 THz ($r_{\text{Rayleigh}} = 609 \mu\text{m}$). At 2 THz, the dynamic range value lowers to 36 dB, until reaching its minimal value of 0 dB at 3.2 THz ($r_{\text{Rayleigh}} = 115 \mu\text{m}$).

The maximum amplitude of the pulse in the time domain represents the average reflected power across the dynamic range [58]. It should therefore be noted that terahertz set-ups with a narrower dynamic range (i.e. poorer dynamic range values in the 1–3-THz region) than the system used in this study would lead to a poorer resolution in images produced from the maximum pulse amplitude than the values reported in the above section.

2.6.4 Signal Processing

Terahertz images from features at a specific interface within a stratified object can be obtained by identifying the pulse corresponding to the layer of interest in the reflected waveform, and producing a map of its amplitude throughout the scanned area [59–62]. For a layer of ink applied on a support 80 to 300 μm thick, two to three interfaces may be considered, depending on the thickness of the ink layer and penetration of the ink into the fibres of the support: the air-ink interface, the ink-front surface of the support interface and the back surface of the support-air interface. Depending on the thickness of the ink and support layers with respect to the bandwidth of the incident terahertz pulse, pulses reflected from each of these interfaces may overlap. This overlap would be greatly influenced by the optical properties of the ink and the support and would change the amplitude, time position and shape of the peaks observed in the collected waveform. Using the amplitude of a specific peak in the collected waveform without further signal processing to objectively compare the contrast in terahertz images from different ink dots on different supports would therefore be unsound. The comparison should rather be based on imaging data obtained from waveforms processed to minimise the influence of the overlap between pulses reflected from different interfaces.

Deconvolution attempts to retrieve the theoretical impulse response function and thereby improves the separation between the two (or three) pulses [11, 47]. Deconvolution of the collected waveform therefore enables the retrieval of a signal in which the peak amplitudes are little or no longer affected by the pulses overlap. Terahertz images from ink dots presented in this study were therefore produced using the maximum amplitude of the deconvolved signals, rather than the raw signals, calculated for each scanning position (i.e. each pixel) using the built-in deconvolution tool in the TVL Imaging Suite software (TeraView Ltd). This deconvolution tool uses a double Gaussian filter [47]. The same high-frequency and low-frequency cut-offs were consistently used for all scanned ink dots.

2.6.5 Normalisation of Images and Calculation of Contrast

To objectively compare the image contrast between ink dots, the maximum amplitude value $M(i,j)$ of the deconvolved waveform obtained for each pixel (corresponding to the amplitude of the pulse reflected at the air-front surface of the sheet interface) was normalised using the following formula:

$$M_{\text{stand}}(i,j) = \frac{M_{\text{reduced}}(i,j)}{\text{sSUB}_{\text{avg}}} \quad (1)$$

with M_{reduced} as the cropped image to obtain an image size of $3.6 \times 3.6 \text{ mm}^2$ (area within the user-defined red rectangle in Fig. 1) and sSUB_{avg} the average of the pixel values corresponding to the support (area outside of the user-defined yellow circle in Fig. 1). All grey scales were set from 1 (black) to 1.5 (white): while the support always appears dark in the normalised terahertz images, the more reflective the ink dot is in the terahertz region, the brighter it appears in the image.

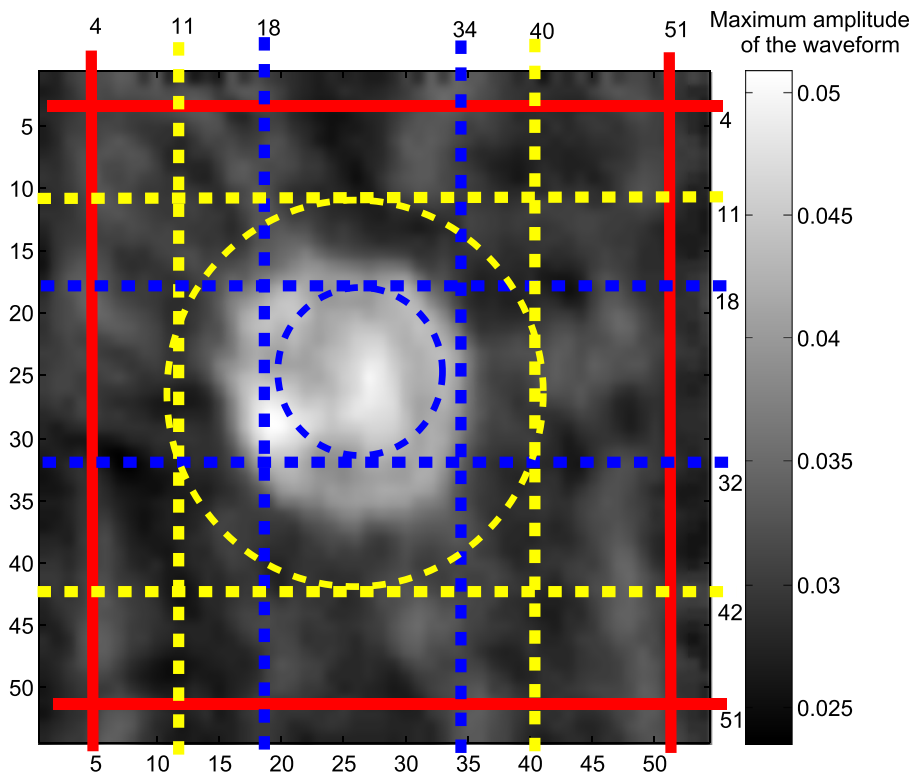


Fig. 1 Terahertz image of a vermilion ink dot obtained, after deconvolution, from the maximum amplitude of the reflected waveform. The image is overlaid by a schematic diagram describing the boundaries used for the image normalisation and calculation of the contrast value C . Red lines correspond to the cropped area to obtain a $3.6 \times 3.6\text{-mm}^2$ image (corresponding here to 48 pixels, with one pixel corresponding to a $75\text{-}\mu\text{m}$ scanning step). All pixels within the blue circle, the diameter of which corresponds to the minimum of the distances between blue horizontal or vertical lines, correspond to the region of interest within the ink dot. All pixels outside the yellow circle, the diameter of which corresponds to the maximum of the distances between yellow horizontal or vertical lines, correspond to the region of interest of the support

To enable the simultaneous comparison of contrast between terahertz images of 43 different ink dots on supports, a contrast value C was calculated from all terahertz images and summarised in Table 2:

$$C = \frac{sROI_{avg}}{sSUB_{avg}} \quad (2)$$

with $sROI_{avg}$ as the average of the pixel values corresponding to the inked area (area inside the user-defined blue circle in Fig. 1). Regions of interest excluded the vicinity of the ink dot edges, as upon drying, some ink droplets created ring-shaped solid residues inducing unpredictable scattering of the terahertz beam. Large differences of reflectivity between the ink and the support lead to large C values.

2.6.6 Step Size

When using a scanning step size (i.e. a pixel size) half the width of the ink dot, there is always a pixel which is fully inscribed within the dot and faithfully renders the signal from the ink, showing maximum contrast, regardless of the position of the dot with regards to the pixel grid. Using a scanning step size below half the width of the ink line does not, in theory, improve the contrast of the dot. Yet a smaller step size may enable to render the width and shape of the ink dot more accurately and to appreciate small variations in peak amplitude due to changes in topography and/or concentrations of the ink within the dot. To illustrate this concept, terahertz images of the same vermilion and minium ink dots on rag paper were acquired with two different step sizes: 75 and 250 μm (images provided in Online Resource 2). The difference in calculated contrast values C between images with a different step size was 0.05 for both the vermilion and the minium ink dot ($C_{\text{vermilion}, 75 \mu\text{m}} = 1.48$, $C_{\text{vermilion}, 250 \mu\text{m}} = 1.43$, $C_{\text{minium}, 75 \mu\text{m}} = 1.37$, $C_{\text{minium}, 250 \mu\text{m}} = 1.32$). In comparison, the difference in image contrast recorded for two dots of vermilion ink applied on the same support and scanned with the same scanning step size can be as large as 0.22. The influence of the step size on the image contrast is therefore not considered to be significant. Still, the rendering of the width, shape and ink distribution of the ink dot is poor in images acquired with a 250- μm step size. To recover it as accurately as possible, images can be processed by increasing their sampling rate through the use of an interpolation filter. The Matlab function *imresize* has been used for this purpose [48]. The resulting images are shown in Online Resource 2 and have the same number of pixels as the images acquired with a 75- μm step size. The width, shape and ink distribution of the ink dot in these processed images closely resemble those observed in images acquired with a 75- μm step size, although a blur can still be observed. Additionally, the change in contrast value compared to unprocessed images does not exceed 0.02 ($C_{\text{vermilion}, 250 \mu\text{m}, \text{image resizing}} = 1.43$ and $C_{\text{minium}, 250 \mu\text{m}, \text{image resizing}} = 1.30$): image resizing has therefore no significant influence on the image contrast value. Reducing the step size to 75 μm therefore seems to add little to the information already enclosed in images acquired with a 250- μm step size. From the discussion in the above sections on image resolution and signal to noise, we learned that the diameter of the 2D PSF of the terahertz pulsed beam can be estimated to be above 230 μm (i.e. $r_{\text{Rayleigh}, 3.2 \text{ THz}} * 2$). It follows that with a step size of 250 μm , the amount of information on the interaction between the incident pulse and the sample that can be collected is maximised. With a step size of 75 μm , this information is therefore oversampled. This is why image resizing yields similar results as reducing the step size but with a slightly poorer

rendering of the width, shape and ink distribution of the dot. In this study, scans were acquired with a step size between 75 and 150 μm , with no further image resizing. However, for future terahertz studies of documents, to avoid acquiring large size data and to reduce data acquisition time while rendering the width, shape and ink distribution of the ink inscriptions as accurately as possible, a large scanning step size (provided this scanning step size is less than half the width of the ink line) combined with image resizing can therefore be used.

3 Results and Discussion

3.1 Categorisation Based on the Constant Refractive Index

Constant refractive indices from all pellets and supports are displayed in Table 1. Interestingly, there seems to be a correlation between the refractive index values from pigments and binders with their electrical conductivity properties. According to Coutaz et al. [63], dry dielectrics have a low refractive index (below ≈ 2), intrinsic semiconductors have a larger refractive index (between ≈ 2 and 5), while metals have a very large refractive index. Coutaz et al. further remind that metals in the terahertz range can be considered as perfect electrical conductors, i.e. as having electrons unbound to atoms and free to move under the influence of the applied terahertz electromagnetic field. The Drude model can therefore be applied to describe the electronic and optical properties of metals in the terahertz range. This model helps to evidence that metals absorb much more strongly terahertz radiation than they reflect it. Paper and parchment are dielectrics. Arabic gum constituted of carbohydrates (as in paper) and rabbit skin glue constituted of collagen (as in parchment) can also be considered as dry dielectrics. It is therefore not surprising that these materials exhibit the lowest refractive indices in Table 1. Graphite is a semi-metal (i.e. a conductor along the plane of each carbon sheet) [64]. It is thus

Table 1 Constant refractive indices of different pigments, binders and supports. For each material, three spectroscopic pellets were produced and analysed once. The average value and standard deviation of the constant refractive index are reported in this table. Data in roman, italics and bold correspond, respectively, to materials unlikely, likely and very likely to lead to a well-contrasted terahertz image when applied on support

Sample	n_0
Arabic gum	1.74 ± 0.06
Bistre	1.77 ± 0.01
<i>Bone black</i>	2.17 ± 0.01
Carbon black	No transmitted signal
<i>Genuine ivory black</i>	1.83 ± 0.07
Graphite	No transmitted signal
<i>Iron gall ink</i>	2.01 ± 0.05
Lamp black	No transmitted signal
<i>Lapis lazuli</i>	2.02 ± 0.02
<i>Malachite</i>	2.20 ± 0.14
<i>Minium</i>	1.95 ± 0.01
Rabbit skin glue	1.46 ± 0.06
Sepia	1.70 ± 0.01
Verdigris	1.45 ± 0.02
<i>Vermilion</i>	1.93 ± 0.10
Original rag paper	1.29 ± 0.02
Additionally sized rag paper	1.33 ± 0.06
Parchment	1.71 ± 0.02

expected to have a high refractive index and an even higher absorption coefficient: this would explain why no signal is transmitted through a pellet of graphite with a 10% mass concentration. Carbon black comes from the incomplete combustion of gaseous hydrocarbons and, as pointed out by Hauptman et al. [65], consists of graphite-like crystalline and amorphous carbon domains. The authors also indicate that the major factor that determines the electrical conductivity of carbon black particles is the graphitic character giving delocalized π -electrons [65]. Lamp black, ivory black and bone black also result from the combustion of carbonaceous materials [20] and may also contain some graphitic and some amorphous carbon. Yet, ivory black and bone black also contain inorganic calcium phosphate and calcium carbonate: they would therefore have a lower mass fraction of carbon than graphite, carbon black or lamp black within the 10% mass concentration pellet. As a result, they are likely to contain a lower mass fraction of graphitic carbon too, and would be less conductive, so would have a lower refractive index and absorption coefficient. This would explain why a transmitted signal is measured through pellets of ivory black and bone black. Bistre contains aromatic carbons connected to functional groups containing hydrogen and oxygen atoms, resulting from the charred lignin [34]: there is therefore an even smaller mass fraction of pure carbon to potentially form graphitic structures in bistre pellets than in ivory black or bone black pellets. It could therefore be assumed that bistre is less conductive than ivory black and bone black, which would be why it shows a lower refractive index in Table 1. Minium (Pb_3O_4) and vermilion (or cinnabar, $\alpha\text{-HgS}$) can be considered as semiconductors [66, 67]. Minium and vermilion therefore have a larger refractive index (close to 2 as reported in Table 1) than dielectrics such as Arabic gum or rabbit skin glue.

The image contrast of ink inscriptions on support depends on the difference between the refractive index of the ink and the refractive index of the support. Inks with a refractive index similar to the refractive index of the surrounding support are unlikely to give high image contrast and are presented as roman in Table 1. Inks with a refractive index much larger than the refractive index of the surrounding support are very likely to give a good image contrast and are presented as bold in Table 1. All other inks with refractive indices larger than the refractive indices of supports, but still allowing the signal transmitted through the pellet to be detected, are likely to give a well-contrasted terahertz image. This categorization of pigments and inks based solely on the semi-quantitative comparison of their refractive indices with those from supports may appear subjective and needs to be confirmed by observing the contrast in terahertz images from ink inscriptions made with those same pigments and binders and applied on those same supports.

3.2 Images from Peak Amplitude

The first thing to observe when reading Table 2 and comparing Figs. 2, 3 and 4 is that lamp black ink dots systematically lead to a very good image contrast. Additionally, for a given support, sepia, bistre and verdigris ink dots systematically belong to the category of lower image contrast values. The other ink dots show image contrasts between these two extremes. Categories listed in Table 1 are therefore confirmed when looking at image contrast, and conclusions about the likelihood of different pigments leading to a good image contrast are reinforced.

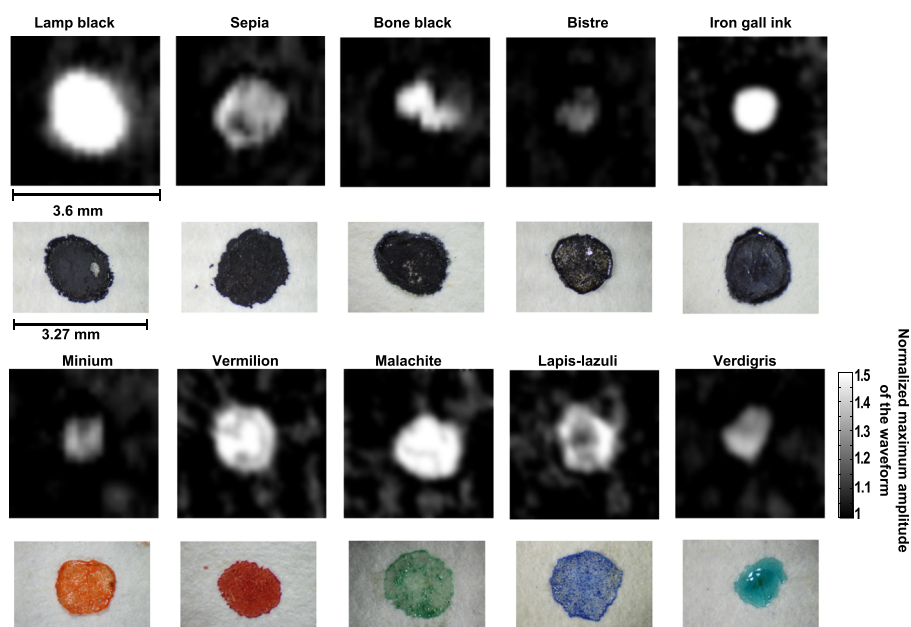
Additionally, comparing Figs. 2 and 4, it can be observed that image contrasts are systematically poorer when imaging ink dots on parchment compared to rag paper sheets. This can be explained by the fact that the refractive index of the parchment sheet used as a

Table 2 Contrast values C extracted from terahertz images of different ink dots on different supports. Images acquired in reflection mode and produced by mapping the amplitude of the first collected peak of the deconvolved waveform for each pixel

	Original rag paper	Additionally sized rag paper	Parchment
Lamp black	2.51	2.10	1.38
Sepia	1.28	1.16	0.94
Bone black	1.39	1.49	1.16
Bistre	1.19 ± 0.04	1.13 ± 0.06	1.04 ± 0.05
Iron gall ink	1.64	1.47	0.74
Minium	1.37 (a) and 1.25 (b)	1.13	1.05
Vermilion	1.41 ± 0.17	1.15 ± 0.05	1.13 ± 0.11
Malachite	1.51	1.22	1.06
Lapis	1.33	1.29	1.01
Verdigris	1.32	1.22	1.02

Values reported for bistre and vermilion correspond to the average C value for three different scanned ink dots, together with their associated standard deviation. Values reported for minium on original rag paper correspond to C values for (a) the non-uniformly distributed ink dot displayed in Online Resource 2 and further discussed in Figs. 8 and 9 and (b) the uniformly distributed ink dot displayed in Fig. 2

support is larger (1.71 ± 0.02) than the refractive indices of rag paper sheets, with or without additional sizing (1.29 ± 0.02 and 1.33 ± 0.06). The differences in refractive index values for inks and the surrounding support are therefore reduced in the case of parchment, and it follows that the differences in the amplitude of the pulse reflected at the air-front surface of the sheet interface with or without ink present are lower for parchment. As for rag paper sheets with and without additional sizing, no significant difference in the constant refractive

**Fig. 2** Photographs of ink dots on original rag paper and the associated terahertz images acquired in reflection mode. Terahertz images were produced using the maximum amplitude of the reflected waveforms after deconvolution. All images were normalised and are displayed on the same grey scale

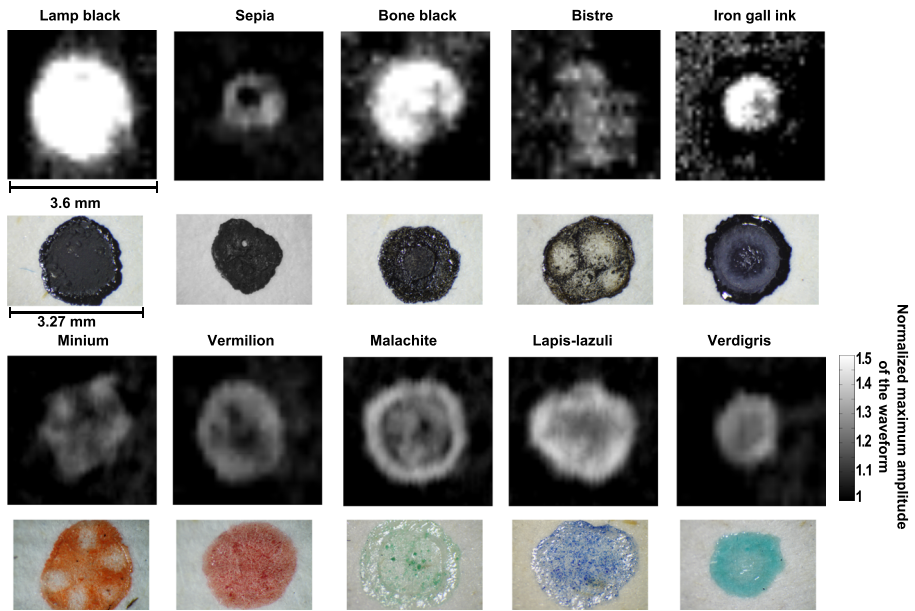


Fig. 3 Photographs of ink dots on additionally sized rag paper and the associated terahertz images acquired in reflection mode. Terahertz images were produced using the maximum amplitude of the reflected waveforms after deconvolution. All images were normalised and are displayed on the same grey scale

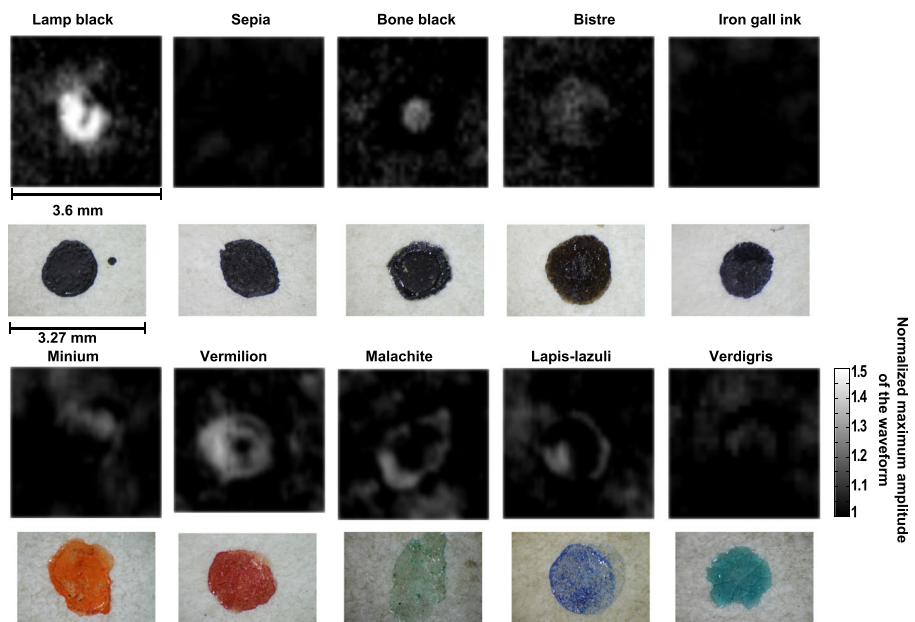


Fig. 4 Photographs of ink dots on parchment and the associated terahertz images acquired in reflection mode. Terahertz images were produced using the maximum amplitude of the reflected waveforms after deconvolution. All images were normalised and are displayed on the same grey scale

index is observed. However, ink dots applied on the additionally sized rag paper sheet tend to have a broader diameter than ink dots on the original rag paper sheets, while the applied volume of the ink remained the same (2 mL). This implies that sizing modifies the surface properties of paper in such a way that the angle of contact between the ink and the sheet is increased. With the exception of bone black (which seems to have undergone no or little spread in the presence of sizing), contrast values tend to be poorer for the rag paper sheet with additional sizing than for the original rag paper sheet. Pigment concentration therefore plays a role in the image contrast: a larger concentration of pigment on the paper surface leads to an increased amplitude of the reflected waveform. This is further confirmed by the variations of contrast within an ink dot (mostly) coinciding with variations of pigment concentration, as clearly illustrated by the minium ink dot on additionally sized paper in Fig. 3.

When not caused by variations in concentration, variations in contrast within an ink dot can arise from scattering. The configuration of the optics within the imaging set-up is such that the illuminating source is at 30° from the normal to the surface, and a detector is also placed at 30° from the normal to the surface. An edge can be considered as a diffuse reflector: the beam would be reflected in all directions, but the intensity of the radiation (and therefore also the amplitude of the electric field) reflected from this edge would be larger in the direction of the normal to the surface, according to Lambert's cosine law. A schematic diagram describing the distribution of the intensity of the electromagnetic field reflected from the surface of an ink dot is available in Online Resource 3. As a result, the amplitude of the electric field collected from an edge at an oblique angle would be reduced compared to a specular reflection. This would appear dark in the terahertz image produced from the amplitude of the pulse reflected from the front surface of the sheet. This phenomenon is particularly observable for bone black on any support, or for iron gall ink and bistre on the original rag paper: due to scattering, the coffee-stain-like ring left upon drying appears darker than the centre of the ink dot, even though it may represent an area with larger pigment concentration.

3.3 Repeatability

To acknowledge the difference between variations of C values inherent to minor variations in shape, concentration or topography of the dots, and variations of C values caused by the different refractive properties of different inks, standard deviation values have been calculated after applying the same ink suspensions of vermilion (colour pigment likely leading to a good image contrast) and bistre (black pigment unlikely to lead to a good image contrast) in triplicate droplets on the same support and imaging them with terahertz TD imaging. In Fig. 5 and Table 2, it can be observed that all bistre ink dots systematically lead to a poor image contrast, while the contrast of vermilion ink dots can vary greatly between triplicates on the same support, likely due to scattering and distribution of ink concentration. Similarly, the local accumulation of minium pigment in the ink dot displayed in Online Resource 2 seems to be at the origin of the larger contrast value reported in Table 2, compared to the contrast value reported for the minium ink dot with an apparent uniform pigment distribution, displayed in Fig. 2. This is the reason why the influence of the refractive index from pigments and supports on the image contrast can only be qualitatively, and not quantitatively, assessed using trends and categories rather than comparing the contrast from ink dots pairwise.

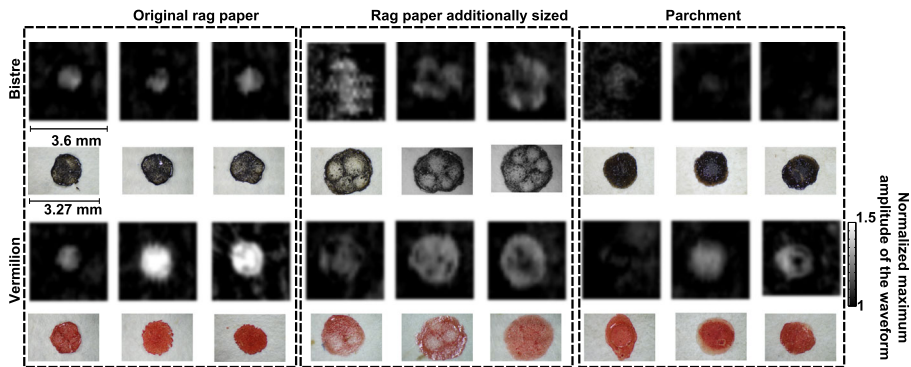


Fig. 5 Photographs of three different bistre and vermilion ink dots applied on rag paper, additionally sized rag paper and parchment, and the associated terahertz images acquired in reflection mode. Terahertz images were produced using the maximum amplitude of the reflected waveforms after deconvolution. All images were normalised and are displayed on the same grey scale

3.4 Extraction of Spectral Information from Ink Inscriptions

While the constant refractive index is a good indicator of the reflective properties of a material with respect to the terahertz pulse, which encompasses all frequencies, materials may exhibit a

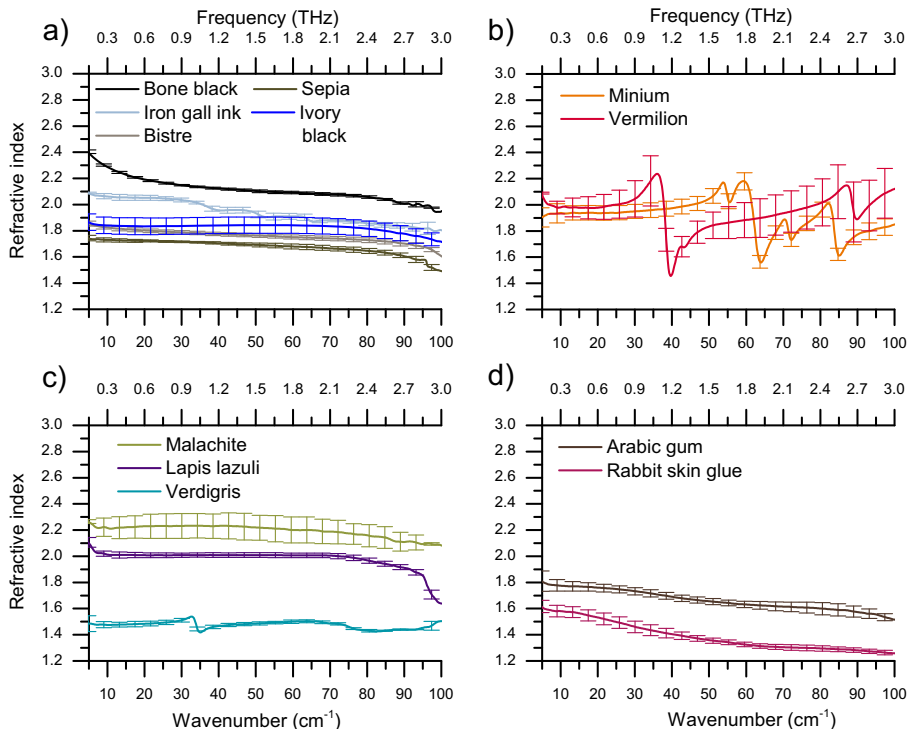


Fig. 6 **a–d** Frequency-dependent refractive indices from different dry pigments, inks and binders, mixed with HDPE powder and pressed in duplicate or triplicate pellets. The *lines* represent the average of data from triplicate or duplicate pellets, and the *error bars* represent standard deviations

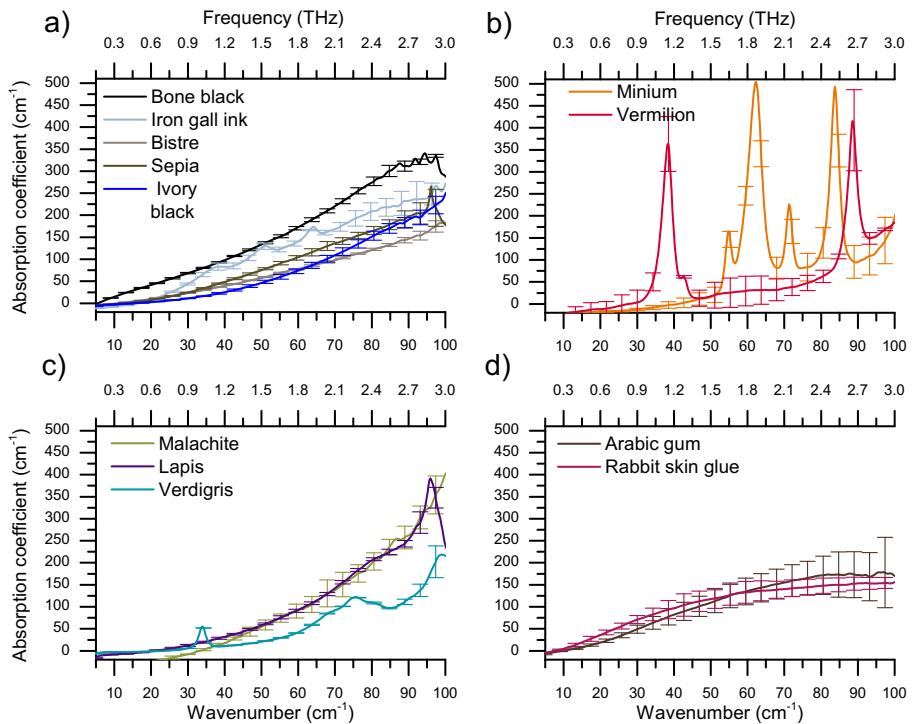


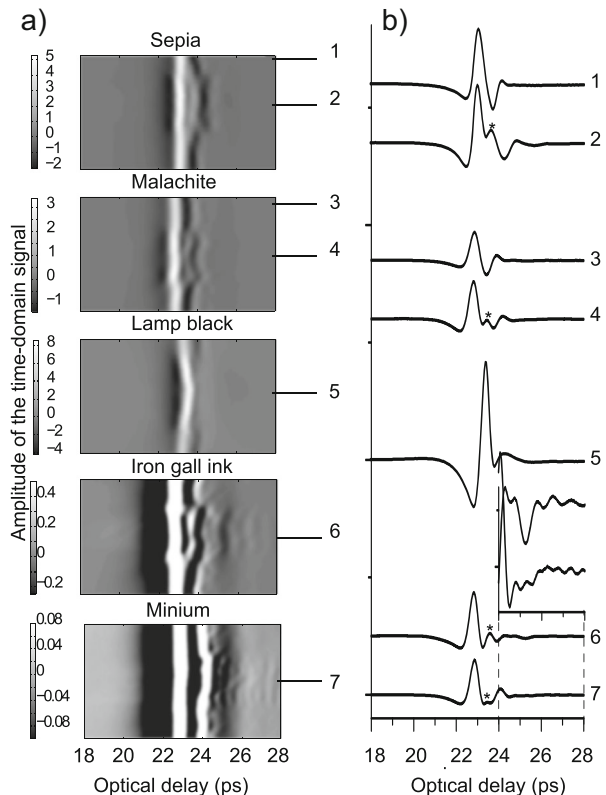
Fig. 7 **a–d** Frequency-dependent absorption coefficients from different dry pigments, inks and binders, mixed with HDPE powder and pressed in duplicate or triplicate pellets. The *lines* represent the average of data from triplicate or duplicate pellets, and the *error bars* represent standard deviations

dispersive behaviour in the terahertz range. The frequency-dependent refractive index from pigments and dry inks is displayed in Fig. 6. Additionally, their frequency-dependent absorption coefficients are displayed in Fig. 7. The first thing to observe is that the samples of vermilion, minium, verdigris and dry iron gall ink exhibit a reproducible spectral signature between 0.15 and 3 THz. Figure 7 shows that vermilion and minium exhibit intense and narrow absorption bands at the resonance frequencies 38, 42 and 89 cm^{-1} (1.15, 1.27 and 2.66 THz) and 55, 62, 71 and 84 cm^{-1} (1.64, 1.86, 2.14 and 2.51 THz), respectively. Iron gall ink shows three broad absorption bands at 38, 51 and 64 cm^{-1} (1.14, 1.53 and 1.92 THz), and verdigris shows a sharp absorption band at 34 cm^{-1} (1.02 THz) and a broad absorption band at 76 cm^{-1} (2.27 THz). The specimens of minium, vermilion, verdigris and iron gall ink studied here are therefore crystalline materials with a resonant behaviour in the terahertz range 0.1–3 THz. It should also be noted that the absorption features from vermilion (cinnabar) and minium are in agreement with the spectral signatures of the supplies from Zecchi, Florence, analysed with a different terahertz TD set-up in transmission and available on the online database thzdb.org [17]. The supply of verdigris from Zecchi, analysed using Fourier-transform infrared spectrometer in the range 0.6–13 THz and available in the same online database also displays absorption bands at ≈ 1 and 2.25 THz, but the signal to noise ratio is poor below 5 THz. The resonant behaviour translates into an “anomalous” dispersion in refractive index values: the refractive index is larger than the resonant frequency at lower frequencies, and it is lower at larger frequencies.

If pigments and inks exhibit larger absorption or reflection properties in specific spectral regions, an improved image contrast of the inscriptions could potentially be obtained using the amplitude of the FD signal in these specific spectral regions only. Besides, these features are characteristic of a specific pigment or ink, so the frequency-dependent behaviour of the contrast in the FD image could reveal the nature of the ink or pigment. In other words, using the mathematical tool of Fourier-transform to convert TD waveforms into FD spectra, and benefiting from the strong resonant behaviour of some pigments and inks in the terahertz region, terahertz TD imaging could potentially be used as a chemical imaging tool.

Figure 8 shows the TD B-scans and raw waveforms from ink dots on rag paper. An interesting phenomenon is observed in the B-scans from iron gall ink and minium ink dots on rag paper and also in the TD B-scan from vermilion shown in Fig. 9: a “ringing” effect is observed in the tail of the TD waveform reflected from an inked area, which translates into “ripples” observed after the back surface of the paper sheet on the TD B-scan, as if additional layers were present under the paper back surface. This phenomenon is systematically observed in all iron gall, minium and vermilion ink dots on rag paper and parchment sheets, although far less pronounced for iron gall ink dots, and is never observed for ink dots made of other pigments. These ink dots are the only pigments with characteristic spectral signatures in the terahertz range (with the exception of verdigris with a weak spectral feature at a low frequency and a very broad

Fig. 8 **a** Terahertz TD B-scans along a line across an ink dot on rag paper (sepia, malachite, lamp black, iron gall ink or minium mixed with Arabic gum) and **b** associated raw TD waveforms extracted from pixel positions indicated with numbers. The *inset* in **b** shows the tail of the waveforms collected in positions 6 and 7 on a magnified scale. The *star* indicates the presence of an additional TD feature due to the presence of an ink dot at the surface of the sheet



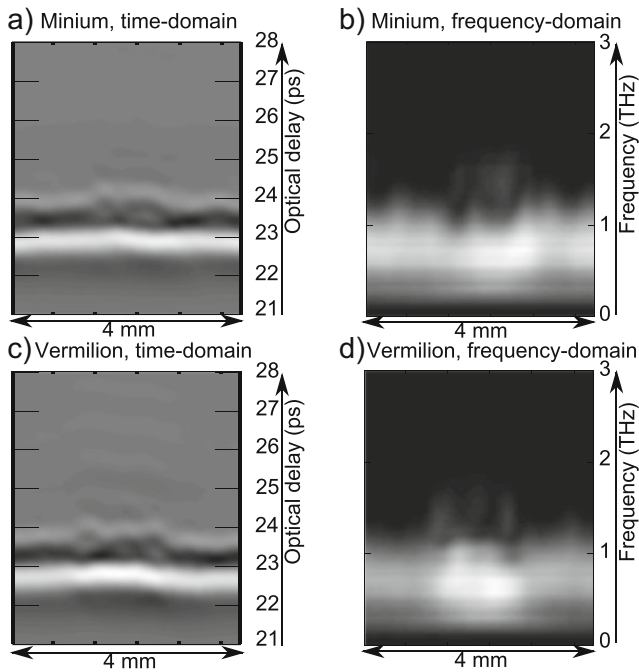


Fig. 9 TD and FD terahertz B-scans along a line crossing the ink dot. **a, b** Minium ink dot on rag paper. **c, d** Vermilion ink dot on rag paper

feature at a high frequency). More specifically, the “ringing” and “rippling” phenomena are very clearly observed for the two pigments with very strong and narrow absorption bands: minium and vermilion. This ringing effect is also clearly observed in the tail of the waveforms transmitted through spectroscopic pellets of minium or vermilion mixed with HDPE (data not shown here). The frequency of this “ringing” effect observed in the signal transmitted through spectroscopic pellets was 1.7 THz for pellets of minium and 1.2 THz for pellets of vermilion. These frequencies correspond to spectral regions of high absorption in the absorption coefficients of minium and vermilion shown in Fig. 7. The frequency of the “ringing” effect observed in waveforms reflected from minium and vermilion ink dots on rag paper and parchment sheets showed similar values (between 1.7 and 1.8 THz for minium and 1–1.1 THz for vermilion). It can therefore be assumed that there is a frequency-dependent absorption of the terahertz pulse transmitted through the thin layer of ink applied on paper and reflected at the ink-paper interface and that this frequency dependence is characteristic of the vibrational properties in the region 0.1–3 THz of the pigment. Unfortunately, this “ringing” effect has a weak amplitude and is significantly attenuated or lost when imaging the same ink dots in “verso” configuration, i.e. with the incident pulse impinging on the air-support interface, then support-ink, and finally ink-air. This is due to the fact that the reflected pulse that has been partially absorbed by the ink layer on the back surface travels through a layer of support before reaching the detector in reflection mode, and is therefore attenuated. This “ringing” effect may also be attenuated or lost during the deconvolution process, depending on the filter function used [47].

Besides, the presence of this “ringing” effect in TD waveforms does not seem to significantly influence the shape of the FD spectra from ink dots on support obtained from the Fourier-transform of the reflected TD waveforms: it is difficult to observe a clear decrease in amplitude at either 1.7 or 1.2 THz for the pixels corresponding to minium and vermilion ink dots, respectively, in the FD B-scans shown in Fig. 9b, d. However, topographic features and morphology of the layered structure have a clear influence on the FD signals [49]. The presence of an ink layer introduces an increase in the optical path, which changes the time delay between the signals reflected from the front and back surfaces of the sheet. This phenomenon and the presence of a new interface create the appearance of an additional layer between the front and back surface of the sheet, i.e. the grey patches observed between 23 and 24 ps in the TD B-scans for pixels within the inked area (Fig. 9a, c). Since pigment distribution within the ink dot, thickness of the ink layer and penetration of the ink into paper fibres are not perfectly uniform across an ink dot, two grey patches are observed for the minium ink dot and three for the vermilion ink dot, instead of a continuous grey layer. The exact same grey patches are observed in the FD B-scans within a frequency range where the rest of the pixels display no (or low) signal (Fig. 9b, d). The FD response therefore seems to be largely influenced by the number of layers in the document and their thickness. To summarise, when imaging a document with terahertz TD technology in reflection mode, the terahertz pulse can undergo dispersive absorption when transmitted through a thin layer of ink on support before it is reflected on the ink-support interface, but it seems that morphological modifications due to the presence of this ink layer influence more the FD response of the document than the resonant behaviour of the materials present within the document.

4 Conclusions

The influence of the difference in refractive index between the ink and the support underneath and around it on the image contrast produced from the amplitude of the pulse reflected at the air-front surface of the sheet interface has been demonstrated. Sepia ink and inks produced with bistre or verdigris mixed with a solution of Arabic gum or rabbit skin glue are unlikely to lead to well-contrasted images. However, bone black, ivory black, iron gall ink, malachite, lapis lazuli, minium and vermilion are likely to lead to well-contrasted images, while lamp black, carbon black and graphite give the best imaging results. This categorisation enables to provide recommendations to paper and archives conservators on occasions when terahertz TD imaging in reflection mode is a valuable depth-imaging technology to uncover text within a document. For inks containing sepia, bistre or verdigris, the use of terahertz TD imaging in reflection mode is therefore not recommended, unless the set-up has a very large signal to noise ratio and can probe very low differences in reflective properties. In addition to simply adapt the grey scale or brightness scale of the image, image processing algorithms can further be used to improve the image contrast of inscriptions and for character recognition, such as k-means clustering algorithms [68].

While the dispersive behaviour of iron gall ink, vermilion and minium can be probed in reflection mode, the sample morphology (e.g. changes in the number and thickness of layers) appears to influence more the FD response of the document than the resonant

behaviour of the materials present within the document. As a result, the use of terahertz TD imaging in reflection mode as a chemical imaging tool for the improvement of image contrast using specific spectral windows or for material identification is currently limited. If there is prior knowledge about the document, such as an estimate of the number and thickness range of the investigated layers, simulating the signal that has propagated through the document could help in recovering the spectral signature from the different materials present within it [69]. However, the number of layers in a sealed archival document is rarely known. Besides, such approach remains very complex and its complexity increases with the number of layers in the document. Terahertz TD imaging in transmission mode may be preferred for chemical imaging, as the FD response would be less influenced by the sample morphology. Yet the set-up needs to show an improved signal to noise ratio in the frequency range of interest in order to detect small changes in the dispersion of the signal caused by the ink [16, 70]. Alternatively, a continuous wave imaging set-up tuned at the frequency where the material of interest shows strong absorption or reflection may be used to improve the contrast of any feature containing this material (e.g. terahertz sources tuned at 1.15, 1.27 or 2.66 THz to improve the contrast of vermilion ink lines and tuned at 1.64, 1.86, 2.14 or 2.51 THz for minium ink lines) [71–73]. Compact terahertz quantum cascade laser sources or Gunn diodes can be used for this purpose and can further be combined with tomography to recover some depth information.

To further inform conservators of the usefulness of terahertz TD imaging to uncover texts in documents, the terahertz optical properties of synthetic dyes and pigments found in documents from the last two centuries may also be studied, such as aniline, nigrosine, methylene blue and Prussian blue. Finally, this study combined theory, spectroscopy and imaging to provide recommendations on occasions when terahertz TD imaging is a valuable depth-imaging technology to uncover text within a document. To strengthen the reliability of these recommendations, the first step would be to objectively assess if the refractive index and absorption coefficient presented in this study faithfully render the optical properties of the associated pigment. To do so, it would be recommended to perform further analysis using a large number of spectroscopic pellets, but also using different supplies of the same pigment, as well as using different set-ups (round robin test). The accuracy of these optical coefficients would also need to be improved by taking pellet porosity and particle size into account, through the combined use of helium gas pycnometry and effective medium theory [74]. Once a more complete and accurate database of optical coefficients from ink and support materials is designed, each terahertz researcher would need to refine the contrast categories described above to the signal to noise imaging performances of his or her terahertz set-up, in order to provide customized system-related recommendations.

Acknowledgements The authors would like to thank the partners TeraView, Ltd and the Nationaal Archief. Professor Sir Michael Pepper, UCL Department of Electronic and Electrical Engineering is gratefully acknowledged for his advice. UK Engineering and Physical Research Council Doctoral Training Centre in Virtual Environments, Imaging and Visualisation is acknowledged for the financial support. This work was additionally supported by The Engineering and Physical Sciences Research Council [grant number EP/G037159/1]. The authors are also grateful to Dr. Lik-Ren Tai, UCL Institute for Sustainable Heritage, for his help in acquiring Fourier-transform infrared spectra.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and

reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. The State Secretary of Education Culture and Science: Dutch Public Records Act 1995. (1995).
2. United Kingdom: Public Records Act 1958. (1958).
3. United Kingdom: Freedom of Information Act 2000. (2000).
4. The European Commission: Commission recommendation of 27.10.2011 on the digitisation and online accessibility of cultural material and digital preservation. Official Journal of the European Union. (2011).
5. The National Archives; United Kingdom: Digitisation at The National Archives. (2016), <https://nationalarchives.gov.uk/documents/information-management/digitisation-at-the-national-archives.pdf>. Accessed 31 December 2016
6. Endangered Archives Programme: Guidelines for photographing and scanning archival material. (2015), http://eap.bl.uk/downloads/guidelines_copying.pdf. Accessed 31 December 2016
7. Saffady, W.: Introduction to Automation for Librarians, 4th edn. (American Library Association, 1999).
8. Hu, B.B., Nuss, M.C.: Imaging with terahertz waves. *Opt. Lett.* 20, 1716 (1995).
9. Labaune, J., Jackson, J.B., Pagès-Camagna, S., Duling, I.N., Menu, M., Mourou, G. A.: Papyrus imaging with terahertz time domain spectroscopy. *Appl. Phys. A Mater. Sci. Process.* 100, 607–612 (2010).
10. Wallace, V.P., Macpherson, E., Zeitler, J.A., Reid, C.: Three-dimensional imaging of optically opaque materials using nonionizing terahertz radiation. *J. Opt. Soc. Am. A. Opt. Image Sci. Vis.* 25, 3120–33 (2008).
11. Walker, G.C., Bowen, J.W., Labaune, J., Jackson, J.-B., Hadjiloucas, S., Roberts, J., Mourou, G., Menu, M.: Terahertz deconvolution. *Opt. Express.* 20, 27230–41 (2012).
12. Abraham, E., Younus, A., El Fatimy, A., Delagnes, J.C., Nguéma, E., Mounaix, P.: Broadband terahertz imaging of documents written with lead pencils. *Opt. Commun.* 282, 3104–3107 (2009).
13. Mittleman, D.M., Gupta, M., Neelamani, R., Baraniuk, R.G., Rudd, J. V, Koch, M.: Recent advances in terahertz imaging. *Appl. Phys. B.* 68, 1085–1094 (1999).
14. Mounaix, P., Younus, a., Delagnes, J.C., Abraham, E., Canioni, L., Fabre, M.: Spectroscopy and terahertz imaging for sigillography applications. *J. Eur. Opt. Soc. Rapid Publ.* 6, 11002 (2011).
15. Redo-Sanchez, A., Heshmat, B., Aghasi, A., Naqvi, S., Zhang, M., Romberg, J., Raskar, R.: Terahertz time-gated spectral imaging for content extraction through layered structures. *Nat. Commun.* 7, 12665:1–7 (2016).
16. Fukunaga, K., Ogawa, Y., Hayashi, S., Hosako, I.: Application of terahertz spectroscopy for character recognition in a medieval manuscript. *IEICE Electron. Express.* 5, 223–228 (2008).
17. RIKEN and NICT: Terahertz database. <http://thzdb.org/>. Accessed 31 December 2016
18. Bardón, T., May, R.K., Taday, P.F., Strlič, M.: Systematic study of terahertz time-domain spectra of historically informed black inks. *Analyst.* 138, 4859–4869 (2013).
19. Bat-Yehouda, M.Z.: Les encres noires au Moyen Age (jusqu'à 1600). (Éditions du Centre National de la Recherche Scientifique, 1983).
20. Eastaugh, N., Walsh, V., Chaplin, R., Siddall, R.: *Pigment Compendium, A Dictionary and Optical Microscopy of Historical Pigments.* (Elsevier Ltd., 2008).
21. Orna, M.V., in *Archaeological Chemistry VIII*, ed. By R. A. Armitage, J. H. Burton (American Chemical Society, 2013), p. 3
22. Van Strydonck, M., Boudin, M., De Mulder, G.: the Carbon Origin of Structural Carbonate in Bone Apatite of Cremated Bones. *Radiocarbon.* 52, 578–586 (2010).
23. Derrick, M.R., Stulik, D., Landry, M., J.: *Infrared Spectroscopy in Conservation Science.* (The Getty Conservation Institute, 1999).
24. Kendix, E.L., Prati, S., Mazzeo, R., Joseph, E., Sciutto, G., Fagnano, C.: Far Infrared Spectroscopy in the Field of Cultural Heritage. *e-PRESERVATIONScience.* 7, 8–13 (2010).
25. Kendix, E.L.: *Transmission and Reflection (ATR) Far-Infrared Spectroscopy Applied in the Analysis of Cultural Heritage Materials*, (2009). [Dissertation thesis], Alma Mater Studiorum Università di Bologna. Dottorato di ricerca in Science for Conservation (EPISCON), 22 Ciclo. doi:[10.6092/unibo/amsdottorato/2266](https://doi.org/10.6092/unibo/amsdottorato/2266)
26. Vigouroux, J.P., Husson, E., Calvarin, G., Dao, N.Q.: Etude par spectroscopie vibrationnelle des oxydes Pb3O4, SnPb2O4 et SnPb(Pb2O4)2. *Spectrochim. Acta Part A Mol. Spectrosc.* 38, 393–398 (1982).
27. The National Institute of Standards and Technology, The NIST Chemistry Webbook, NIST Standard Reference Database Number 69 (2016), <http://webbook.nist.gov/chemistry/>. Accessed 31 December 2016

28. Guo, D.: Mechanism of Dissolution of High Valence Lead Oxides in Water Under Depleting Chlorine Conditions. (2014). [Master thesis] University of Western Ontario - Electronic Thesis and Dissertation Repository. Paper 1948. <http://ir.lib.uwo.ca/etd/1948/>. Accessed 31 December 2016
29. Roy, R., White, W.B.: Infrared spectra-crystal structure correlations: II. Comparison of simple polymorphic minerals. *Am. Mineral.* 49, 1670–1687 (1964).
30. Bacci, M., Cucci, C., Del Federico, E., Ienco, A., Jerschow, A., Newman, J.M., Picollo, M.: An integrated spectroscopic approach for the identification of what distinguishes Afghan lapis lazuli from others. *Vib. Spectrosc.* 49, 80–83 (2009).
31. Miliani, C., Daveri, A., Brunetti, B.G., Sgamellotti, A.: CO₂ entrapment in natural ultramarine blue. *Chem. Phys. Lett.* 466, 148–151 (2008).
32. Museum of Fine Arts, Boston, CAMEO: Conservation and Art Materials Encyclopedia Online. http://cameo.mfa.org/wiki/Main_Page. Accessed 31 December 2016
33. Lafuente B., Downs R. T., Yang H., Stone N. The power of databases: the RRUFF project, in *Highlights in Mineralogical Crystallography*, ed. By T. Armbruster and R. M. Danisi (W. De Gruyter, Berlin, 2015) pp 1–30
34. Sharma, R.K., Wooten, J.B., Baliga, V.L., Lin, X., Chan, W.G., Hajaligol, M.R.: Characterization of chars from pyrolysis of lignin. *Fuel* 83, 1469–1482 (2004).
35. Wells, A. F.: Malachite: re-examination of crystal structure. *Acta Crystallogr.* 4, 200–204 (1951).
36. Van Niekerk, J.N., Schoening, F.R.L, X-Ray Evidence for Metal-to-Metal Bonds in Cupric and Chromous Acetate. *Nature*. 171, 36–37 (1953).
37. Espinosa-Andrews, H., Sandoval-Castilla, O., Vázquez-Torres, H., Vernon-Carter, E.J., Lobato-Calleros, C.: Determination of the gum Arabic-chitosan interactions by Fourier Transform Infrared Spectroscopy and characterization of the microstructure and rheological features of their coacervates. *Carbohydr. Polym.* 79, 541–546 (2010).
38. Bardon, T., May, R.K., Taday, P.F., Strlič, M.: Influence of Particle Size on Optical Constants From Pellets Measured With Terahertz Pulsed Spectroscopy. *IEEE Trans. Terahertz Sci. Technol.* 6, 408–413 (2016).
39. Anderson, K.J.: Paper and Other Writing Materials : Part II. *MRS Bull.* 37–38 (1991).
40. Hunter, D.: *Papermaking: the history and technique of an ancient craft.* (Dover Publications, Inc., New York, 1978)
41. Henry, W. et al., in *Paper Conservation Catalog.* (American Institute for Conservation Book and Paper Group, Washington D.C., 1988). http://cool.conservation-us.org/coolaic/sg/bpg/pc/18_parchment.pdf. Accessed 31 December 2016
42. Roberts, M.T., Etherington, D.: *Bookbinding and the Conservation of books: A Dictionary of Descriptive Terminology* (Library of Congress, Washington, D.C., 1982), <http://cool.conservation-us.org/don/tn/tn6.html>. Accessed 31 December 2016
43. Petukhova, T.: A History of Fish Glue as an Artist's Material: Applications in Paper and Parchment Artifacts. *B. Pap. Gr. Annu.* 19, (2000).
44. Reed, R.: *Ancient Skins, Parchments and Leathers.* (Seminar Press, 1972).
45. Kolar, J., Strlic, M., *Iron Gall Inks: On Manufacture, Characterisation, Degradation and Stabilisation.* (National and University Library, Ljubljana, 2006)
46. Kniffin, G.P., Schecklman, S., Chen, J., Henry, S.C., Zurk, L.M., Pejcinovic, B.: Measurement and Modeling of Terahertz Spectral Signatures from Layered Material. 7687, 1–11 (2010).
47. Zeitler, J.A., Shen, Y.-C., in *Terahertz Spectroscopy and Imaging*, ed. By Peiponen, K.-E., Zeitler, J.A., and Kuwata-Gonokami, M. (Springer, 2013) pp. 451–489
48. MATLAB: version R2008a. The MathWorks Inc., Natick, Massachusetts (2008).
49. Bardon, T., May, R.K., Taday, P.F., Strlič, M.: Contrast in terahertz images of archives—Part II: Influence of topographic features. *J. Infrared, Millimeter, Terahertz Waves.* doi:10.1007/s10762-016-0352-z
50. Deegan, R.D., Bakajin, O., Dupont, T.F., Huber, G., Nagel, S.R., Witten, T.: Capillary flow as the cause of ring stains from dried liquid drops. *Nature*. 389, 827–829 (1997).
51. Hecht, E., *Optics*, 4th edn. (Addison-Wesley, 2002)
52. den Dekker, A.J., van den Bos, A.: Resolution: a survey. *J. Opt. Soc. Am. A.* 14, 547 (1997).
53. Sun, J.H., Gallacher, J.G., Brussaard, G.J.H., Lemos, N., Issac, R., Huang, Z.X., Dias, J.M., Jaroszynski, D.: Electro-optic measurement of terahertz pulse energy distribution. *Rev. Sci. Instrum.* 80, 4–7 (2009).
54. Jepsen, P.U., Jacobsen, R.H., Keiding, S.R.: Generation and detection of terahertz pulses from biased semiconductor antennas. *J. Opt. Soc. Am. B.* 13, 2424 (1996).
55. Baron Rayleigh. *Investigations in optics, with special reference to the spectroscope.* In *Scientific papers* (Cambridge University Press, 1899), p. 420
56. Withayachumnankul, W., Naftaly, M.: Fundamentals of Measurement in Terahertz Time-Domain Spectroscopy. *J. Infrared Millim. TerahertzWaves.* 35, 610–637 (2014).

57. Jepsen, P.U., Fischer, B.M.: Dynamic range in terahertz time-domain transmission and reflection spectroscopy. *Opt. Lett.* 30, 29–31 (2005).
58. Fitzgerald, A.J., Berry, E., Miles, R.E., Zinovev, N.N., Smith, M. a, Chamberlain, J.M.: Evaluation of image quality in terahertz pulsed imaging using test objects. *Phys. Med. Biol.* 47, 3865–73 (2002).
59. Koch Dandolo, C.L., Filtenborg, T., Skou-Hansen, J., Jepsen, P.U.: Analysis of a seventeenth-century panel painting by reflection terahertz time-domain imaging (THz-TDI): contribution of ultrafast optics to museum collections inspection. *Appl. Phys. A.* (2015) doi:10.1007/s00339-015-9303-0
60. Anastasi, R.F., Madaras, E.I.: Terahertz NDE For Under Paint Corrosion. *NDT.net*, 11, 7 (2006), <http://www.ndt.net/article/v11n07/papers/05.pdf>. Accessed 31 December 2016
61. Seco-Martorell, C., López-Domínguez, V., Arauz-Garofalo, G., Redo-Sanchez, A., Palacios, J., Tejada, J.: Goya's artwork imaging with Terahertz waves. *Opt. Express*. 21, 17800 (2013).
62. Abraham, E., Younus, A., Delagnes, J.C., Mounaix, P.: Terahertz-pulse imaging for non-destructive analysis of layered art paintings. In: 35th International Conference on Infrared Millimeter and Terahertz Waves (IRMMW-THz), 2010. pp. 2–3 (2010).
63. Coutaz, J.-L., Boquet, R., Breuil, N., Chussey, L., Crozat, P., Demaison, J., Duvillaret, L., Gallot, G., Garet, F., Lampin, J.-F., Lippens, D., Mangeney, J., Mounaix, P., Mouret, G., Roux, J.-F., *Optoélectronique térahertz*. (EDP Sciences, Les Ulis, 2008)
64. Pierson, H.O.: *Handbook of Carbon, Graphite, Diamonds and Fullerenes: Processing, Properties and Applications* (Noyes Publications, 2012)
65. Hauptman, N., Vesel, A., Ivanovski, V., Gunde, M.K.: Electrical conductivity of carbon black pigments. *Dyes and Pigments*, 95, 1–7 (2012).
66. Droessler, L.M.: *Lead Oxides for Photovoltaics*, (2014), [Thesis], University of Oxford. <http://www-solar.materials.ox.ac.uk/uploads/images/theses/thesis%20Laura%20Droessler%20final.pdf>. Accessed 31 December 2016
67. Berger, L.I.: *Semiconductor Materials*. (CRC Press, 1996).
68. Ayech, M.W., Ziou, D.: Segmentation of Terahertz imaging using k-means clustering based on ranked set sampling. *Expert Syst. Appl.* 42, 2959–2974 (2015).
69. Mechelen, J.L.M. Van, Kuzmenko, A.B., Merbold, H.: Stratified dispersive model for material characterization using THz time-domain spectroscopy. *Opt. Lett.* 39, 3853–3856 (2014).
70. Miyamaru, F., Takeda, M.W., Suzuki, T., Otani, C.: Highly sensitive surface plasmon terahertz imaging with planar plasmonic crystals. *Opt. Express*. 15, 14804–9 (2007).
71. Fukunaga, K., Sekine, N., Hosako, I., Oda, N., Yoneyama, H., Sudoh, T.: Real-time terahertz imaging for art conservation science. *J. Eur. Opt. Soc. Rapid Publ.* 3, 1–4 (2008).
72. Hosako, I., Sekine, N., Patrashin, M., Saito, S., Fukunaga, K., Kasai, Y., Baron, P., Seta, T., Mendrok, J., Ochiai, S., Yasuda, H.: At the Dawn of a New Era in Terahertz Technology. *Proc. IEEE*. 95, 1611–1623 (2007).
73. Fukunaga, K., Hosako, I., Ogawa, Y., Hayashi, S.: THz Spectroscopy for analysis of paintings. 2007 Annu. Rep. - Conf. Electr. Insul. Dielectr. Phenom. 715–718 (2007).
74. Parrott, E.P.J., Zeitler, J.A., Gladden, L.F.: Accurate determination of optical coefficients from chemical samples using terahertz time-domain spectroscopy and effective medium theory. *Opt. Lett.* 34, 3722–3724 (2009).